Unsupervised K-means Feature Learning for Gesture Recognition with Conductive Fur

Anonymous Author(s)
Affiliation
Address
email

Abstract

Humans engage in many sophisticated forms of emotional communication, one of which occurs through touch. In the past, this emotional capacity clearly separated humans from machines. But as recent advances in artificial intelligence put the ability to perceive and express emotions through touch within reach of computers, we must ask - how is it that humans so adeptly access emotion through touch, and is this something computers can do? As our group explores this question in the context of emotional touch between a person and a furry social robot, we require sensing able to capture and recognize touch gesture types. To this end, we describe a new type of touch sensor based on conductive fur, which measures changing current as the conductive threads in the fur connect and disconnect during touch interaction. From a data set of these time-series electrical current curves for a set of three key gestures, features are learned with unsupervised k-means clustering. These features are then classified using multinomial logistic regression. Cross-validation of the classifier's performance for a 7-participant data set shows promise for this approach to gesture recognition.

1 Introduction

The human brain is not purely rational; rather it carries out a complex combination of thinking and feeling. Picard [1] argues that therefore, a truly natural symbiosis between people and machines cannot exist without harnessing emotion. Early work in emotional computing has raised a range of controversial questions about the possible roles of emotion in computers, whether for artificial perception, expression, or even possession of emotion. What is clear is that the design of emotionally intelligent haptic experiences offers exciting and important possibilities. Touch-based social robots have been used for empathic communication, and are capable of providing emotional support and companionship. Affective touch is especially important for the development and well-being of the young, the old, the ill and the troubled. Thus there are many valuable social and healthcare-related applications, including rehabilitation, education, treatment of cognitive disorders, and assistance for people with special needs [2, 3, 4].

Current haptic affective systems, which rely largely on force and electric field sensors, are not yet able to classify gestures adequately even if used in combination. This suggests the need for an additional channel of information. In the present research, we describe the design of a new fur-based touch sensor based on above-surface hand motion information (Figure 1), inspired by Buechley's stroke sensor [5] . We extract time-series hand motion information from this sensor, use unsupervised k-means clustering to learn features in the data, and apply multinomial logistic regression to classify gesture. Preliminary results suggests this design could contribute to gesture recognition.

Figure 1: Conductive fur sensor.

2 Related Work

2.1 Touch-Sensitive Social Robots

Huggable, PARO, Aibo and Probo are some of the best-known examples of affective robots that are sensitive to touch [6, 4, 7, 8]. These projects are largely focused around force sensors such as Force Sensitive Resistors (FSRs), and capacitive sensors. While these approaches are promising, the projects are still in early stages of gesture recognition, and current results suggest that neither force nor capacitive sensing is likely to have the sensing scope needed to differentiate gesture. It is therefore of interest to investigate alternate sensor types that could improve recognition accuracy by providing a different channel of information for affective touch.

The goal of this work is to investigate such an alternative channel to contribute to the gesture recognition capabilities of another touch-sensitive affective robot, the Haptic Creature [9]. An animal-like but deliberately non-representational robot, the Haptic Creature senses the world through touch alone, with a focus on identifying human emotional states from touch gestures. The eventual goal of this work is to improve gesture recognition by fusing our sensor's output with the Creature's other sensors.

2.2 Gesture Recognition Technologies in Touch-Sensitive Systems

The use of machine learning for touch gesture recognition in affective systems is in early stages. The designers behind both Huggable and PARO have experimented with supervised neural networks using feature-based sensor data [6, 4]. The Haptic Creature team has also made use of features, with an eventual probabilistic structure in mind [10].

One approach is the use of learning schemes for data mining of time series. To our knowledge, time-series specific learning is unexplored for gesture recognition, a surprising gap given the time-dependent nature of gestures.

Therefore, this work explores feature learning from time-series gesture data. Based on Coates, et. al. [11], we use unsupervised k-means clustering to extract features from our electrical current sequences, which are then classified with multinomial logistic regression.

3 Sensor Design

Before describing our recognition technique, we outline the basic design of our sensor setup, and the data it produces.

Our physical design is based on the observation that during a touch interaction between a human and a furry animal, the hand disturbs the configuration of the animal's fur, with an arguably distinctive pattern. We are interested in capturing physical changes in the fur for visibility into the gesture space.

Figure 2: Buechley's conductive thread stroke sensor (left) [5], our conductive fur touch and gesture sensor (right).
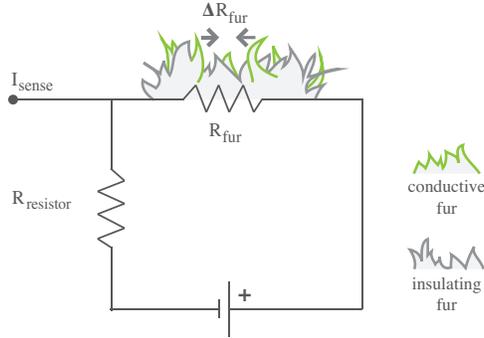
Figure 3: Circuit for our design. Touches change the fur configuration and consequently the net fur resistance, R_{fur}. The resulting fluctuating current=f(time) is sampled at 144 Hz (I_{sense}).

Three key gestures are selected from Yohanan's touch dictionary [12]: *stroke*, *scratch* and *light touch*. [12] defines these gestures as following: *stroke*: moving one's hand gently over the fur, often repeatedly, *scratch*: rubbing the fur with one's fingernails, *light touch*: touching the fur with light finger movements. These gestures are chosen on the basis of crucial affective content [10], inadequate differentiation by existing sensor technology, and a potentially good match to the fur-based sensor.

We are inspired by Buechley's design for a low-tech binary stroke sensor that responds to a *stroke* gesture [5]. In the sensor concept which we have adopted from [5] (Figure 2), a stroking motion brushes the vertically-sewn conductive threads together. When a pair of adjacent threads do not touch, they present infinite resistance to the circuit, and a finite resistance when they do touch. This circuit is effectively made of many resistors connected in parallel; its total resistance drops as more connections are made, and hence measurable current increases (Figure 3).

We build upon this idea in several ways, described in detail in Flagg, et. al. [13]. In summary: first, we sew conductive threads into a sample of the thick fur that is used in the Haptic Creature to create realism and visual, tactile attractiveness. Second, rather than sampling a single "stroke" or "no stroke" state, we sample current over time (I(t)). Third, using I(t) also allows us to position the threads more densely, because we are no longer restricted to maintaining a broken circuit when the threads are not being stroked, which improves touch-sensitive coverage of the fur. Finally, we make use of two layers of different lengths, enriching the data to be more sensitive to touch types that interact with different positions in the fur (i.e., roots vs. top of the fur). See Figure 2 for a visual comparison, and [13] for a detailed description.

4 Analysis

We begin our analysis with a data set made of 210 2-second samples of *stroke*, *scratch* and *light touch*. Data was collected from 7 participants outside the project, each contributing 10 examples for each gesture.

We apply Coates, et. al.'s method for classification based on unsupervised feature learning [11], adapted from image data to our 1-dimensional time-series data, and implemented in Python. Specif-

ically, we use k-means to cluster random subsequences, or *shapelets*, from our training sequences, then express a given data point in terms of how close its shapelets are to the k clusters. The concatenation of distances from each extracted shapelet to each cluster is the resulting feature vector. A regularized logistic regression model is then trained on these features for classification.

Following the algorithmic structure presented in [11], the below steps transform a data sequence to a learned feature representation:

1) Extract random shapelets from the unlabeled training sequences.

2) Learn a feature-mapping using k-means clustering.

We then have a feature mapping and a set of labeled training sequences which can be used for feature extraction and classification:

1) Extract features from equally-spaced shapelets covering each input sequence.

2) Train a logistic regression classifier on these features.

We briefly discuss the structure of these components in our implementation:

Sampling random shapelets

Our first step is to sample m random shapelets from the training set, each of size w. These shapelets are put into a matrix of random shapelets $X = \{x^{(1)}, ..., x^{(m)}\}$, where $x^{(i)}$ belongs to R^w.

Unsupervised feature learning with k-means

Unsupervised k-means clustering is used to learn features of the data. The matrix X of randomly sampled shapelets is grouped into k clusters. Then, given these k learned centroids $c^{(k)}$, we can define the following sparse, non-linear feature mapping:

$$f_k(x) = max\{0, \mu(z) - z_k\}$$

where f_k is the kth element of f, $z_k = ||x - c^{(k)}||_2$, and $\mu(z)$ is the mean of the elements of z.

Thus this step outputs a function $f : R^w \to R^k$ mapping an input sequence to a new feature vector based on k learned centroids.

Extracting features

We now have a function f that maps a shapelet $x \in R^w$ to a new feature vector $y = f(x) \in R^k$. This feature extractor can be applied to our data sequences for training and classification.

Specifically, we extract equally-spaced shapelets of size w from a data sequence, where the space s between the starting point of any given shapelet and the next is referred to as the *stride*. Thus we can represent an input sequence as a list of shapelets, each of which is mapped to its corresponding feature vector. These individual shapelet feature vectors are concatenated to form the complete feature vector F for the entire sequence, where $F \in R^{k*m}$. This is our new representation of the data that will be used as input for classification.

Classification

Finally, given our $(m * k)$-dimensional feature vectors, we apply a standard multinomial logistic regression classifier with L2 regularization.

Parameters

Cross-validation is used to determine the regularization parameter, as well as the optimal shapelet size w, stride s, the number of clusters k, and the number of random patches to extract. Results follow in the next section.

5 Results

We split our 210 gesture samples into 180 training cases and 30 test cases. After training, our most successful logistic regression solver classified the test set with 83.33% accuracy. This performance was achieved with the following parameter values: a shapelet size of 36, a stride of 37, 16 k-means
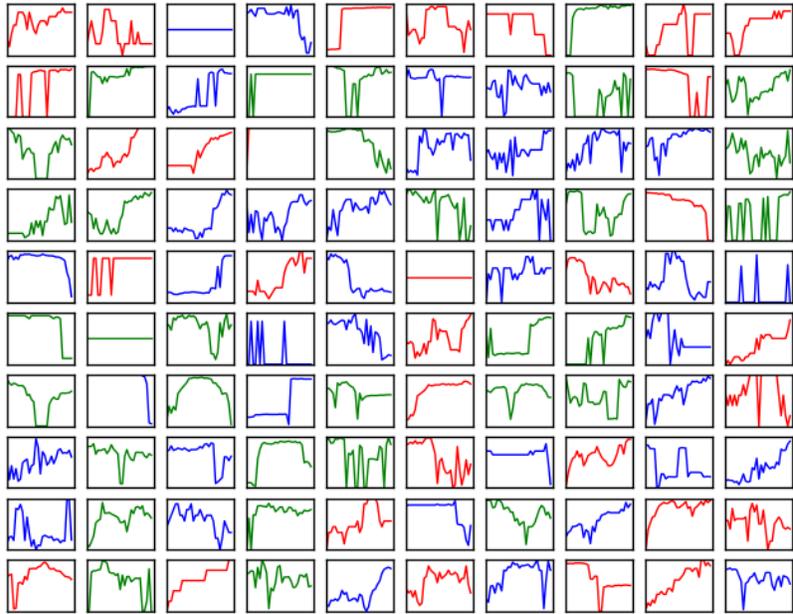
Figure 4: Features learned from unsupervised k-means clustering, colored by class.

clusters, 40,000 random shapelets, and a regularization parameter of 0.1. Figure 4 shows the features clustered with k-means.

6 Discussion and Future Work

Our classifier performance of 83.33% is decent for this early stage in the project, especially given that this is a relatively new and unexplored type of data. However, it will not be sufficient for our long-term goals, and there is still much work to be done to improve it. First, we observe from much experimentation that performance can vary widely for the same choice of parameters. This is due to the randomness present in both the initial shapelet extraction, and the k-means cluster initialization. To counter this we suggest choosing a large number of random shapelets, so the space of possible shapelets is better covered. (Note that in our most successful model, we used 40,000 shapelets.) Next, to deal with the randomness inherent in k-means initialization, we suggest running k-means several times during cross-validation, and choosing the cluster configuration that performs best on the test set. Of course, this will involve splitting the data into training, test, and validation sets, and then measuring the chosen model's ultimate performance with the validation set.

We mentioned that a large number of random shapelets helped stabilize performance. We also noticed that using a relatively small number of clusters for k-means also improved the model, because it discouraged overfitting. Setting a large shapelet size also helped describe the trend of the data, rather than capturing small details in noisy readings.

Another observation we made is that contrary to performance in [11], our data was much better classified *without* preprocessing such as whitening and normalizing. It is not clear exactly why this is, but our intuition is that for our type of data, absolute electrical current values are important, because a strong identifying feature of different gestures is that they physically connect different numbers of conductive hairs in the fur, which affects the overall current flowing through the circuit. If the data is normalized and whitened, then this absolute information is lost. More experimentation will be necessary to confirm this.

To further improve results, we will in future try dynamic time warping on the shapelets. This method is considered state-of-the-art for classifying time-series data [14], but does not seem to have been explored yet for gesture recognition. We could also experiment with smoothing, because visualizations of our data show a lot of noise in the signals. It is also possible that using Euclidean distance as a similarity measure is not the best way to compare shapelets, so we could try other measures. We could also try other unsupervised feature learning methods such as clustering with Gaussians, or spectral clustering. Also, we used logistic regression in this work, but we can try experimenting with other classifiers. Finally, we plan to eventually incorporate data from force sensors to augment our conductive fur readings.

If successful, this work will be integrated into the Haptic Creature to improve gesture recognition. Better gesture recognition in the Creature will provide a better understanding of emotion, which will allow for more intelligent emotional interaction. We hope in this way to contribute to the therapeutic power of emotion-aware furry social robots.

7   Acknowledgements

8   References

[1] Picard, Rosalind. Affective Computing. MIT Media Laboratory; 20 Ames St., Cambridge, MA 02139 (1997).

[2] Okamura, Allison, Mataric, Maja J., and Christensen, Henrik I. "Medical and Health-Care Robotics", IEEE Robotics & Automation Magazine, September 27-37 (2010).

[3] Dautenhahn, Kerstin. I could be you - the phenomenological dimension of social understanding, Cybernetics and Systems Journal, 28(5), 417-453 (1997).

[4] Shibata, T., Inoue, K., and Irie, R. Emotional Robot for Intelligent System: Artificial Emotional Creature Project. In Proceedings of IIZUKA (43-48) (2006).

[5] Buechley, L. Instructable Stroke Sensor. http://www.instructables.com/id/Stroke- Sensor/, May 2011.

[6] Stiehl, W. & Breazeal, C. Design of a Therapeutic Robotic Companion for Relational, Affective Touch. In Proceedings of Fourteenth IEEE Workshop on Robot and Human Interactive Communication (Ro-Man-05), Nashville, TN. 408-415. Best paper Award. (2005).

[7] Friedman, Batya, Kahn, Peter H. Jr., Hagman, Jennifer. Hardware companions?: what online AIBO discussion forums reveal about the human-robotic relationship. CHI: 273-280 (2003).

[8] Goris, K., Saldien, J., Vanderniepen, Innes, Lefeber, D. The Huggable Robot Probo, a Multi-disciplinary Research Platform. Eurobot 2008 Conference, Heidelberg, Germany. (2008),

[9] Yohanan, Steve. and MacLean, Karon E. The Haptic Creature Project: Social Human- Robot Interaction through Affective Touch. In Proceedings of The Reign of Katz and Dogz, 2nd AISB Symp on the Role of Virtual Creatures in a Computerised Society (AISB '08), Aberdeen, UK, 7-11 (2008).

[10] Chang, J., MacLean, K., Yohanan, S. (2010), Gesture Recognition in the Haptic Creature. In Proceedings of the 2010 International Conference on Haptics: Generating and Perceiving Tangible Sensations, Part I. (2010).

[11] Coates, Adam, Honglak, Lee, and Ng, Andrew. An Analysis of Single-Layer Networks in Unsupervised Feature Learning. In AISTATS 14 (2011).

[12] Yohanan, Steve and MacLean, Karon E. The Role of Affective Touch in Human-Robot Interaction: Human Intent and Expectations in Touching the Haptic Creature. International Journal of Social Robotics (SORO), Special Issue on Expectations, Intentions, and Actions (accepted August 2011).

[13] Flagg, Anna, Tam, Diane, MacLean, Karon and Flagg, Robert. Conductive Fur Sensing for a Gesture-Aware Furry Robot. In Proceedings, IEEE Haptics Symposium, March 2012 (accepted November 2011).

[14] Xing, Zhengzheng, Peo, Jian, Keogh, Eamonn. A brief survey on sequence classification. SIGKDD Explorations 12(1): 40-48 (2010).